

University of Groningen

Integriteit en duurzaamheid in het digitale tijdperk

Oldenburg, Beau

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Oldenburg, B. (2015). *Integriteit en duurzaamheid in het digitale tijdperk: Het bewaren, delen, hergebruiken en eocumenteren van digitale onderzoeksdata door sociologen in Nederland*. DANS - Data Archiving and Networking Services.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Integriteit en duurzaamheid in het digitale tijdperk

Het bewaren, delen, hergebruiken en documenteren van
digitale onderzoeksdata door sociologen in Nederland

Integriteit en duurzaamheid in het digitale tijdperk

Het bewaren, delen, hergebruiken en documenteren van
digitale onderzoeksdata door sociologen in Nederland

Data Archiving and Networked Services (DANS)

Postbus 93067 | 2509 AB Den Haag

Anna van Saksenlaan 51 | 2593 HW Den Haag

+31 70 349 44 50

info@dans.knaw.nl | dans.knaw.nl



Inhoud

Achtergrond	2
1 Bewaren van digitale onderzoeksdata	3
1.1 Tijdens het onderzoek	3
1.2 Na het onderzoek	4
2 Digitale onderzoeksdata delen	5
2.1 Programmaleiders	6
2.2 Promovendi	7
2.3 Het repliceren en delen van kwalitatieve data: Een bijzondere uitdaging ..	8
3 Digitale onderzoekdata hergebruiken	9
4 Digitale onderzoekdata documenteren	10
4.1 Ondersteuning van de universiteitsbibliotheek	10
Conclusie	11
Literatuur	13

Achtergrond

De digitale revolutie heeft in de afgelopen decennia een sterke invloed gehad op wetenschappelijk onderzoek. Zo is het door de digitalisering van onderzoeksdata mogelijk om grotere hoeveelheden data sneller te analyseren. Deze digitalisering brengt echter ook nieuwe uitdagingen met zich mee. Een van de belangrijkste uitdagingen is het integer en duurzaam omgaan met digitale data (High Level Expert Group on Scientific Data, 2010). In dit onderzoek¹ is verkend hoe sociologen in Nederland met digitale onderzoeksdata omgaan. In open interviews vertelden veertien sociologen - zes programmaleiders² en acht promovendi - van zes verschillende universiteiten (Rijksuniversiteit Groningen, Universiteit van Amsterdam, Radboud Universiteit Nijmegen, Universiteit Utrecht, Vrije Universiteit en Tilburg University) hoe zij en andere onderzoekers in hun vakgroep met digitale onderzoeksdata omgaan. In het bijzonder vertelden zij over het bewaren, delen, hergebruiken en documenteren van digitale onderzoeksdata binnen hun vakgroep.

¹ Dit onderzoek is begin 2015 uitgevoerd door Beau Oldenburg (Rijksuniversiteit Groningen) in opdracht van DANS. Begeleider vanuit DANS was Marion Wittenberg.

² Onder 'programmaleiders' worden onderzoeksdirecteuren en directeuren van onderzoekslijnen verstaan.

1 Bewaren van digitale onderzoeksdata

Een belangrijk aspect van digitale data is de manier waarop de data opgeslagen worden. Vooral als er geen analoge dragers van de data (meer) zijn, kunnen digitale data bijzonder kwetsbaar zijn. Het is cruciaal om digitale data zo op te slaan dat ze niet verloren gaan bij eventuele technische problemen. Uit zowel de interviews met de programmaleiders als de interviews met de promovendi komt een onderscheid tussen het opslaan van digitale data *tijdens* het onderzoek en het opslaan van digitale data *na* het onderzoek naar voren.

1.1 Tijdens het onderzoek

In geen van de vakgroepen die hebben meegewerkt aan dit onderzoek is er beleid als het gaat om het opslaan van digitale data tijdens het onderzoek. De opslag van data wordt tijdens het onderzoek geregeld binnen de onderzoeksgroepen of individuele onderzoekers regelen dit zelf. Alle universiteiten uit dit onderzoek beschikken over een beveiligde digitale omgeving waarin onderzoekers privé mappen kunnen maken en mappen kunnen delen met collega's binnen de universiteit, maar onderzoekers zijn niet verplicht hier gebruik van te maken. De programmaleiders hebben dan ook geen zicht op in hoeverre onderzoekers binnen hun vakgroep de beveiligde digitale omgeving ook daadwerkelijk gebruiken. De beveiligde omgeving wordt door vijf van de acht promovendi genoemd als zij gevraagd worden hoe digitale data binnen hun vakgroep wordt opgeslagen. Twee programmaleiders en twee promovendi antwoorden op de vraag hoe digitale onderzoeksdata worden opgeslagen dat zij en/of hun collega's de data op een harde schijf opslaan. Zes van de geïnterviewde onderzoekers geven aan data op te slaan en uit te wisselen via Dropbox, USB sticks of e-mail. Belangrijkste redenen om data op deze manieren op te slaan zijn eenvoudige toegang vanuit huis en gemakkelijke uitwisseling van data met coauteurs die werkzaam zijn op een andere (internationale) universiteit. Daarnaast geven vier onderzoekers aan dat de data opgeslagen worden op de server van het onderzoeksbureau dat de data verzameld heeft.



1.2 Na het onderzoek

Vijf van de zes programmaleiders geven aan dat zij na ruchtmakende fraudezaken als de Stapel affaire het gevoel hebben dat de wetenschap onder een vergrootglas is komen te liggen. Wetenschappelijke integriteit is volgens hen dan ook een veelbesproken onderwerp in de colleges van bestuur van de universiteiten die meegewerkt hebben aan dit onderzoek. De programmaleiders geven aan dat zij van mening zijn dat de sociologie minder fraudegevoelig is dan bijvoorbeeld de psychologie omdat er vaak met meerdere onderzoekers aan grootschalige datasets gewerkt wordt. Desalniettemin voelt men toch druk van de media, financiers en wetenschappelijke tijdschriften om op een meer transparante manier met de data om te gaan.

In de vakgroepen is veel discussie geweest over op welke manieren fraude het best kan worden voorkomen. In vijf van de zes groepen is recentelijk beleid geïntroduceerd wat de controleerbaarheid van onderzoek moet vergroten. Dit beleid komt er kortweg op neer dat onderzoekers zodra zij een artikel publiceren een *publication package* moeten maken met de syntax en data waarop het artikel gebaseerd is. In drie vakgroepen is dit beleid reeds volledig geïmplementeerd, in twee vakgroepen is men aan het onderzoeken hoe dit beleid het beste geïmplementeerd kan worden. Opmerkelijk is dat dit beleid in twee van de vijf vakgroepen enkel van toepassing lijkt te zijn op promovendi en niet op andere (senior) onderzoekers. Vier onderzoekers geven aan dat het invoeren van deze *publication packages* weliswaar een goede ontwikkeling is, maar dat het ook veel tijd kost om de *publication packages* te maken. Drie van de vijf programmaleiders van vakgroepen waarin de *publication packages* zijn (of worden) ingevoerd merken op dat fraude met onderzoeksdata simpelweg niet volledig uit te bannen is, maar dat de *publication packages* mogelijk wel drempels opwerpen.

Er is variatie in hoe binnen de vakgroepen de *publication packages* worden gecontroleerd. De *publication packages* worden veelal opgeslagen op de eerder beschreven beveiligde digitale omgeving van de universiteit. Een nadeel hiervan is dat anderen buiten de universiteit er geen toegang toe hebben. De controle op de *publication packages* gebeurt dan ook met name intern. In drie vakgroepen wordt steekproefsgewijs gecontroleerd. In deze vakgroepen wordt de syntax van willekeurig geselecteerde artikelen gecontroleerd en wordt bekeken of de resultaten overeenkomen met de resultaten in het gepubliceerde artikel. In de twee overige vakgroepen wordt in eerste instantie uitsluitend bekeken of de syntax en data daadwerkelijk zijn opgeslagen in een *publication package*. In deze vakgroepen wordt alleen gecontroleerd als daar aanleiding toe is. In geen van de vakgroepen is er duidelijk beleid als het gaat om wat er gebeurt als de *publication packages* niet in orde blijken te zijn.

2 Digitale onderzoeksdata delen

De voornaamste reden om data en syntaxen op te slaan in zogenaamde *publication packages* is de replicateerbaarheid van het onderzoek.

Repliceerbaarheid is een van de belangrijkste pijlers van wetenschappelijk onderzoek (VSNU, 2012). De *publication packages* zijn in principe niet bedoeld om data met anderen te delen voor hergebruik. Een manier waarop digitale data gedeeld kunnen worden met anderen is door deze onder te brengen in zogenaamde repositories. Data in repositories kunnen open toegankelijk zijn (*open access*) of toegankelijk zijn voor bepaalde personen (*restricted access*). Het online archiveringssysteem EASY van DANS is een bekende Nederlandse repository waar digitale data van onder andere sociaalwetenschappelijk onderzoek gearchiveerd kunnen worden.

Het duurzaam archiveren van data heeft positieve uitkomsten voor individuele wetenschappers, maar ook voor de wetenschap in het algemeen (Van den Eynden & Bishop, 2014). Als data toegankelijk zijn voor andere onderzoekers kunnen deze onderzoekers de data gebruiken om nieuwe onderzoeksvragen mee te beantwoorden. Door data te delen worden tijd, geld en moeite gespaard en wordt er dus efficiënter met de data omgegaan.

In de afgelopen jaren zijn er diversen partijen geweest die zich hebben ingezet om digitale data zoveel mogelijk voor anderen beschikbaar te maken.

Uitgangspunt hierbij is vaak dat onderzoek dat gefinancierd is met publieke middelen beschouwd moet worden als een publiek goed en daarom ook voor anderen toegankelijk moet zijn (OECD, 2007; High Level Expert Group on Scientific Data, 2010). The Royal Society (2012) schrijft zelfs in haar rapport 'Science as an Open Enterprise' dat het niet delen van data beschouwd kan worden als slechte wetenschap.

In lijn hiermee, verlangen financiers van onderzoek in toenemende mate van onderzoekers dat digitale onderzoeksdata gedeeld worden met anderen. Grote financiers van onderzoek zoals NWO, KNAW en ZonMw hebben als standpunt dat onderzoeksdata die door hen gefinancierd zijn zo toegankelijk mogelijk moeten zijn voor anderen (NWO, 2015).



2.1 Programmaleiders

Alle programmaleiders uit dit onderzoek geven aan bekend te zijn met DANS. Volgens vijf van de zes programmaleiders worden data die in de vakgroep verzameld worden – mits deze hiervoor geschikt zijn – gedeponeerd bij DANS. Vier van de zes programmaleiders voegen hier aan toe DANS een zeer geschikte plek te vinden als het gaat om het archiveren van data. Met name de centrale functie van DANS spreekt hen aan. Zij geven aan hun data liever bij een landelijk archief als DANS te archiveren dan bij een lokaal archief zoals de universiteitsbibliotheek. Volgens alle programmaleiders worden vooral grootschalige kwantitatieve datasets bij DANS ondergebracht. Het deponeren bij DANS is echter niet verplicht: vier programmaleiders geven aan dat dit binnen de onderzoeksgroep of door de individuele onderzoeker besloten wordt. De overige twee programmaleiders geven aan dat onderzoeksdata uit hun vakgroep altijd bij DANS worden gearhiveerd. Of dit in deze vakgroepen ook gecontroleerd wordt, blijft onduidelijk.

De programmaleiders noemen diverse redenen om data met anderen te delen. Repliceerbaarheid en kennisvermeerdering worden het vaakst (vijf keer) genoemd. Vier programmaleiders voegen hieraan toe dat het delen van onderzoeksdata simpelweg deel van hun baan als wetenschapper is. Drie programmaleiders noemen dat data meestal gefinancierd zijn met publieke middelen en zij vinden daarom dat deze gedeeld moeten worden met anderen. Dat het goed is voor de onderzoeker als de data gebruikt worden door anderen wordt drie keer als reden om data te delen genoemd. Ten slotte geven twee programmaleiders aan dat eisen van financiers of wetenschappelijke tijdschriften invloed hebben op het delen van onderzoeksdata.

Het is opmerkelijk dat hoewel vrijwel alle programmaleiders aangeven dat onderzoeksdata in hun vakgroep ondergebracht worden bij DANS dit in de praktijk niet op deze schaal lijkt te gebeuren. Dit is in lijn met eerder onderzoek (Dillo & Doorn, 2011; Van den Eynden & Bishop, 2014) waaruit bleek dat ondanks dat onderzoekers positief tegenover het archiveren en delen van data

lijken te staan een overgrote meerderheid de data niet deponeren bij een archief en de data uitsluitend op hun eigen computer opslaan. Het lijkt erop dat de belangrijkste reden om data niet te archiveren en delen is dat onderzoekers vrezen dat daarmee het onderzoek van degenen die de data hebben verzameld – met name promovendi – in gevaar komt. Vier programmaleiders geven dan ook aan dat de data doorgaans met restricties worden gearchiveerd om de onderzoekers die de data verzameld hebben te beschermen. Zo worden de data voor een bepaalde periode zo gearchiveerd dat deze niet toegankelijk is voor anderen of moeten onderzoekers die de data willen gebruiken eerst een publicatieplan indienen. Slechts één programmaleider geeft aan niet bang te zijn dat anderen er met de data vandoor gaan en onderzoeksdata *open access* te archiveren.

2.2 Promovendi

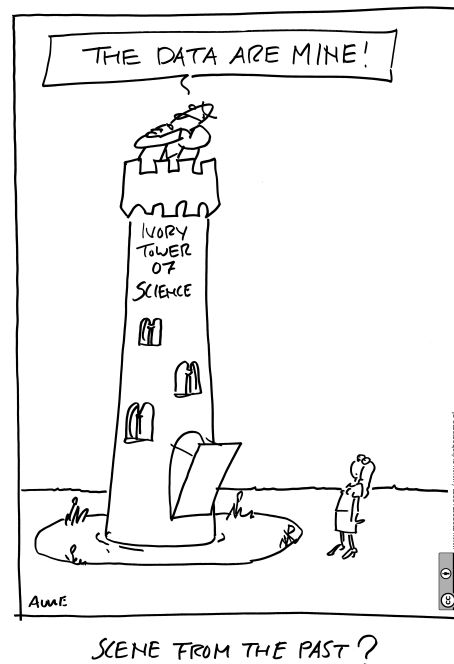
Zes van de acht promovendi verzamelen zelf data. De ene helft van deze groep doet dit binnen een groter onderzoeksproject, de andere helft verzamelt zelfstandig data. Slechts één promovendus weet wat er met de zelf verzamelde data zal gebeuren als het onderzoek afgerond is. Deze promovendus was onderdeel van een groter project waarbij het archiveren van de data eis van de financier was. De overige vijf promovendi geven aan (nog) niet nagedacht te hebben over de toekomst van hun onderzoeksdata.

Vier van de zes promovendi die zelf data verzamelen geven aan dat een reden om data mogelijk met andere onderzoekers te delen, is dat het onderzoek meestal gefinancierd is met publiek geld. Daarnaast werden controleerbaarheid, efficiëntie en eisen van de financier van het onderzoek ieder één keer genoemd als reden om data met anderen te delen. Over het algemeen lijken de promovendi terughoudend als het gaat om het delen van hun data. Hoewel ze aangeven dat ze zich ervan bewust zijn dat dit een onwetenschappelijke houding is, vinden vier van de zes promovendi die zelf data verzamelen het oneerlijk als anderen 'zomaar' gebruik zouden kunnen maken van hun data. Zij geven aan dat dataverzameling veel tijd en moeite kost en dat zij als promovendi hier niet altijd direct credits voor krijgen. Vier promovendi geven aan dat zij het minder erg vinden om hun data te delen met iemand die zij kennen dan met iemand die zij niet kennen. Twee promovendi geven aan dat zij hun data niet vrijblijvend willen delen en dat daar minimaal een co-auteurschap of voetnoot tegenover zou moeten staan.

Deze bevinding sluit aan bij eerder onderzoek (Van den Eynden & Bishop, 2014) waaruit bleek dat onderzoekers vaak bang zijn voor competitie van anderen als ze hun data delen. Door data te delen kunnen anderen mogelijk nog onontdekte interessante resultaten vinden en deze publiceren. Daar komt nog eens bij dat het archiveren tijd, moeite en soms geld kost (Van den Eynden & Bishop, 2014). Vanuit dit perspectief zijn voor individuele onderzoekers de kosten van het delen van data aanzienlijk hoger dan de baten.

Een andere reden om data niet met anderen te delen die uit dit onderzoek naar voren komt, is de angst dat anderen de data mogelijk verkeerd zullen begrijpen. Alle drie de promovendi die zelfstandig data hebben verzameld, geven aan dat ze bang zijn dat anderen hun data wellicht verkeerd zullen interpreteren. Zij geven

aan dat het negatieve gevolgen voor hen kan hebben als anderen over de verkeerd geïnterpreteerde data publiceren.



2.3 Het repliceren en delen van kwalitatieve data: Een bijzondere uitdaging

In het merendeel van de vakgroepen wordt naast kwantitatief onderzoek ook kwalitatief onderzoek gedaan. Zonder dat hier direct naar gevraagd is, merken zes van de geïnterviewde onderzoekers op dat het repliceren en delen van kwalitatieve data een grote uitdaging vormt. Volgens de geïnterviewde onderzoekers is er in kwalitatief onderzoek geen syntax die opgenomen kan worden in een *publication package*. Hierdoor is het lastig één beleid te ontwerpen dat op iedereen van toepassing is. Daarnaast merken de geïnterviewde onderzoekers op dat kwalitatieve onderzoekers, en met name etnografen, vaak vinden dat hun onderzoeksdata niet los van de onderzoeker geïnterpreteerd kunnen worden, wat het delen van data bemoeilijkt. Volgens de geïnterviewde onderzoekers is het archiveren en delen van data een traditie in kwantitatief sociaalwetenschappelijk onderzoek, maar is dit niet het geval als het gaat om kwalitatief onderzoek. De programmaleiders zouden graag advies en ondersteuning krijgen bij het repliceren en delen van kwalitatief onderzoek binnen hun vakgroep.

3 Digitale onderzoekdata hergebruiken

Zoals al eerder genoemd werd, bespaart het hergebruiken van onderzoeksdata tijd, geld, en moeite. In andere woorden: hergebruik gaat verspilling tegen. Tegenwoordig vragen financiers zoals ZonMw onderzoekers dan ook om in hun onderzoeksvoorstel te beargumenteren waarom hun onderzoeksvragen niet met bestaande data beantwoord kunnen worden.

In de vakgroepen die hebben meegewerkt aan dit onderzoek is er geen beleid als het gaat om het hergebruiken van data. De projecten van de promovendi liggen vaak al grotendeels vast: zij hebben zelf weinig invloed op de beslissing om data te hergebruiken of zelf data te verzamelen. De belangrijkste redenen om data te hergebruiken is volgens de geïnterviewde onderzoekers dat data verzamelen veel tijd kost. Deze reden wordt door zeven geïnterviewde onderzoekers genoemd. Met name voor promovendi is het zelf verzamelen van data een grote tijdsinvestering. Drie programmaleiders merken op dat dataverzameling vaak duur is. Met name in grote landenvergelijkende projecten is het qua tijd en geld niet haalbaar om zelf data te verzamelen. Twee geïnterviewde onderzoekers merken op dat het hergebruik van data voorkomt dat respondenten nodeloos belast worden.

Een nadeel van het hergebruiken van data is volgens de geïnterviewde onderzoekers dat je aangewezen bent op de keuzes die anderen hebben gemaakt. Zo kan het zijn dat de data oude of onbetrouwbare instrumenten bevatten of niet goed aansluiten bij de theorie. Maar liefst tien geïnterviewde onderzoekers noemen dit als reden om zelf data te verzamelen. Acht onderzoekers vinden dat het zelf verzamelen van data een waardevolle toevoeging aan de vaardigheden van de promovendi is.

4 Digitale onderzoekdata documenteren

Hoe onderzoekers met de data omgaan, kan beschreven worden in een Data Management Plan (DMP). In een DMP worden alle keuzes en handelingen in het onderzoek gedocumenteerd. Het samenwerkingsverband universiteitsbibliotheken en koninklijke bibliotheek (UKB) definieert een DMP als “een template of checklist voor het beschrijven van de omgang met en het beheer van onderzoeksdata”. Naast dat een DMP de onderzoeker helpt om dieper na te denken over het verzamelen, beschrijven en opslaan van de data, maakt een DMP het onderzoek ook begrijpelijk voor anderen. In een DMP wordt de context van het onderzoek op een dusdanige manier beschreven dat onderzoekers die niet direct bij het onderzoek betrokken waren de data ook kunnen begrijpen en gebruiken. Er zijn diverse soorten templates (voor een overzicht zie <https://wiki.surfnet.nl/display/RD/Datamanagementplannen>) beschikbaar die aanknopingspunten voor het schrijven van een DMP kunnen bieden.

Hoewel Nederlandse financiers van onderzoek in toenemende mate van onderzoekers verwachten dat zij een DMP schrijven bij subsidieaanvragen, lijkt deze term niet bekend onder de geïnterviewde onderzoekers. Slechts twee van de veertien geïnterviewde onderzoekers geven aan bekend te zijn met de term DMP. Na doorvragen, lijken de meeste onderzoekers wel over DMP onderwerpen nagedacht te hebben, maar hebben zij deze nooit opgeschreven in een officieel document. De geïnterviewde onderzoekers geven aan niet vooraf een plan te maken maar over DMP onderwerpen na te denken op het moment dat ze hiermee in aanraking komen. In sommige projecten blijkt de DMP informatie in het codeboek opgenomen te zijn. Er blijkt in geen van de vakgroepen beleid te zijn wat het schrijven van een DMP verplicht.

4.1 Ondersteuning van de universiteitsbibliotheek

Een aantal universiteitsbibliotheken bieden ondersteuning op het gebied van onderzoeksdata aan als dienst. Zo kunnen ze bijvoorbeeld advies geven als het gaat om het maken van een DMP. Zeven geïnterviewde onderzoekers blijken niet bekend te zijn met de ondersteuning die de universiteitsbibliotheek kan bieden. Slechts één geïnterviewde onderzoeker heeft gebruik gemaakt van de ondersteuning van de bibliotheek. Deze onderzoeker geeft aan hier erg tevreden over te zijn. Daarentegen twijfelen drie programmaleiders of de universiteitsbibliotheek de aangewezen organisatie is om advies over dergelijke onderwerpen te geven. Zij benadrukken dat de universiteitsbibliotheek wellicht waardevolle assistentie kan bieden in andere disciplines, maar dat hier binnen de sociologie minder noodzaak voor is. Zij zijn van mening dat onderzoekers geen ondersteuning nodig hebben en dat als zij toch ondersteuning zouden willen zij dat liever zouden krijgen van een landelijke organisatie als DANS.

Conclusie

Uit dit onderzoek blijkt dat er diverse positieve ontwikkelingen zijn als het gaat om integriteit en duurzaamheid van digitale onderzoeksdata. Zo bieden universiteiten onderzoekers een digitale omgeving waar digitale onderzoeksdata veilig opgeslagen kunnen worden. Ook worden onderzoeksresultaten in toenemende mate gecontroleerd middels *publication packages*.

Desalniettemin blijkt dat er ook ruimte is voor meer ontwikkeling. Een groot aantal onderzoekers blijkt hun data tijdens het onderzoek op te slaan op kwetsbare, onbetrouwbare of onveilige media. De voornaamste reden om van deze media gebruik te maken, is eenvoudige toegang vanuit thuis en de mogelijkheid om de data gemakkelijk te delen met andere onderzoekers buiten de universiteit. Mogelijk kunnen de beveiligde digitale omgevingen op de universiteiten zo aangepast worden dat zij makkelijker toegankelijk zijn vanuit thuis en voor onderzoekers die op andere universiteiten werkzaam zijn.

Daarnaast is het belangrijk dat onderzoekers meer bewust worden van de risico's en nadelen van het gebruik van media als USB sticks en Dropbox.

De *publication packages* zijn in vrijwel alle vakgroepen ingevoerd om fraude met onderzoeksresultaten te voorkomen. Vanuit dit oogpunt is het opmerkelijk dat in sommige vakgroepen enkel de promovendi dergelijke *packages* moeten maken. De achterliggende gedachte is waarschijnlijk dat dit beleid het best kan worden ingevoerd onder de jongste generatie wetenschappers. Echter de vraag is of dit de meest effectieve manier is om fraude tegen te gaan. Daarnaast is het in sommige vakgroepen niet duidelijk hoe en door wie de *packages* gecontroleerd worden. Programmaleiders geven aan repliceerbaarheid belangrijk te vinden, maar zijn ook huiverig om een sfeer van achterdocht te creëren. Meestal worden de *publication packages* intern gecontroleerd, maar wellicht kan fraude beter tegengegaan worden door middel van externe controle. In ieder geval lijkt het wenselijk dat vooraf duidelijk is wat de gevolgen zijn als deze *packages* niet in orde zijn, iets wat nu niet het geval is.

In lijn met eerder onderzoek (Dillo & Doorn, 2011; Van den Eynden & Bishop, 2014), lijken onderzoekers welwillend tegenover het delen van data te staan, maar doen zij dit in de praktijk weinig. Een deel van de datasets die volgens de geïnterviewde onderzoekers bij DANS gearchiveerd zijn, blijken dit niet te zijn. Het is onduidelijk waarom dit precies zo is. Het zou kunnen dat dit veroorzaakt wordt door slechte communicatie binnen grotere onderzoeksprojecten. Wat ook een verklaring kan zijn is dat onderzoekers positief tegenover het archiveren van onderzoeksdata in het algemeen staan, maar terughoudend zijn als het om hun eigen data gaat. Zowel onder programmaleiders als onder promovendi blijkt een weerstand om data met anderen te delen. Dataverzameling kost veel tijd, geld en moeite en de onderzoekers vinden het oneerlijk als anderen er met de data vandoor kunnen gaan en ermee kunnen publiceren. Een mogelijke oplossing hiervoor is om binnen de vakgroepen strikter beleid in te voeren als het gaat om het delen van data. Echter dit beleid heeft waarschijnlijk het meest effect als het voor onderzoekers aantrekkelijker wordt gemaakt om hun data te delen met anderen. Wanneer datasets gearchiveerd worden bij een *repository* zoals DANS EASY krijgen de data een uniek identificatienummer (DOI) waarmee deze geciteerd kunnen worden. Deze citaties wegen echter (nog) niet zo zwaar als bijvoorbeeld citaties naar wetenschappelijke artikelen. Een andere mogelijkheid om het delen van data te bevorderen is om het mogelijk te maken om data te publiceren.

Een andere manier om het delen van data te bevorderen is door promovendi hier meer bewust van te maken. Een substantieel deel van de promovendi blijkt niet nagedacht te hebben over de toekomst van hun data. Veel van deze promovendi blijken ook niet bekend te zijn met de diensten van DANS.

Een onderwerp wat naar voren kwam uit de interviews is het delen van kwalitatieve onderzoeksdata. De geïnterviewde onderzoekers gaven aan dat zij onbekend zijn met de mogelijkheden om kwalitatieve data te archiveren en delen en dat zij vaak op bezwaren stuiten bij collega's die kwalitatief onderzoek verrichten. Hoewel, het archiveren van kwalitatieve data anders in zijn werk gaat dan het archiveren van kwantitatieve data is dit wel degelijk mogelijk. DANS heeft in 2009 met het project '*Veteran tapes*' veel kennis opgedaan op het gebied van (secundaire analyse van) kwalitatieve data, zoals het duurzaam archiveren van interviews, het regelen van de juridische randvoorwaarden (bescherming privacy en gebruikslicenties) en het toepassen van geavanceerde ICT-technieken, zoals automatische spraakverwerking en het koppelen van onderzoeksdata aan wetenschappelijke publicaties. (Van Den Berg, Scagliola & Wester, 2010). Dit project heeft geleid tot de archivering van verschillende grote *oral history* collecties in DANS EASY.

Het is opmerkelijk dat vrijwel niemand van de geïnterviewde onderzoekers bekend is met de term Data Management Plan. De onderzoekers blijken gaandeweg het onderzoek wel over deze onderwerpen na te denken, maar zetten deze niet op papier. Vooral als het gaat om zaken als eigenaarschap van de data is het echter raadzaam afspraken van tevoren op papier vast te leggen. Dit kan mogelijke problemen in latere stadia van het onderzoek voorkomen. Recent heeft NWO aangekondigd dat onderzoekers die een NWO subsidie hebben gekregen de kosten voor datamanagement kunnen financieren vanuit deze subsidie (NWO, 2015). Op deze manier probeert NWO goed datamanagement te bevorderen. Ten slotte, blijken onderzoekers weinig bekend met de ondersteuning die universiteitsbibliotheken kunnen bieden als het gaat om integer en duurzaam omgaan met onderzoeksdata. Een deel van de onderzoekers geeft aan dat hier binnen de sociologie weinig noodzaak toe is en dat zij liever ondersteuning zouden ontvangen van DANS dan van de universiteitsbibliotheek.

Literatuur

Dillo, I., & Doorn, P. (2011). The Dutch data landscape in 32 interviews and a survey. Beschikbaar via http://depot.knaw.nl/10090/1/The_Dutch_Datalandscape_DEF.pdf

High Level Expert Group on Scientific Data (2010). Riding the wave: How Europe can gain from the rising tide of data. Beschikbaar via <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>

NWO (2015). NWO-Strategie 2015-2018. Beschikbaar via <http://www.nwo.nl/over-nwo/strategie>

OECD (2007). OECD Principles and guidelines for access to research data from public Funding. Beschikbaar via <http://www.oecd.org/sti/sci-tech/38500813.pdf>

The Royal Society (2012). Science as an open enterprise. Beschikbaar via https://royalsociety.org/~media/royal_society_content/policy/projects/sape/2012-06-20-saoe.pdf

Van Den Berg, Scagliola, S., & Wester, F. (2010). Wat Veteranen Vertellen; verschillende perspectieven op biografische interviews over ervaringen tijdens militaire operaties. Amsterdam; Pallas Publications-AUP.

Van Den Eynden, V., & Bishop, L. (2014). Incentives and motivations for sharing research data, a researcher's perspective. Beschikbaar via https://europe.rd-alliance.org/sites/default/files/report/KE_report-incentives-for-sharing-researchdata.pdf

VSNU (2012). De Nederlandse Gedragscode Wetenschapsbeoefening (herziening 2012). Beschikbaar via [http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Code_wetenschapsbeoefening_2004_\(2012\).pdf](http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Code_wetenschapsbeoefening_2004_(2012).pdf)



Begin 2015 heeft Beau Oldenburg (Rijksuniversiteit Groningen) in opdracht van DANS onderzocht hoe sociologen in Nederland omgaan met hun digitale data, met name op welke manier digitale data worden bewaard, gedeeld, hergebruikt en gedocumenteerd. Neem voor meer informatie over het onderzoek contact op met Beau Oldenburg via b.oldenburg@rug.nl. Neem voor meer informatie over het archiveren of hergebruiken van data via DANS contact op met Marion Wittenberg, via marion.wittenberg@dans.knaw.nl.

Data Archiving and Networked Services (DANS)

DANS bevordert duurzame toegang tot digitale onderzoeksgegevens. Hiertoe stimuleert DANS dat wetenschappelijke onderzoekers gegevens duurzaam archiveren en hergebruiken, bijvoorbeeld via het online archiveringssysteem EASY (easy.dans.knaw.nl) en DataverseNL (dataverse.nl). Tevens biedt DANS met NARCIS (narcis.nl) toegang tot duizenden wetenschappelijke datasets, publicaties en andere onderzoeksinformatie in Nederland. Daarnaast verzorgt het instituut training en consultancy en doet het onderzoek naar duurzame toegang tot digitale informatie. Gedreven door data zorgt DANS er met zijn dienstverlening en deelname in (inter-)nationale projecten en netwerken voor dat de toegang tot digitale onderzoeksgegevens verder verbetert. Kijk op dans.knaw.nl voor meer informatie en contactgegevens.

Data Archiving and Networked Services (DANS)

Postbus 93067 | 2509 AB Den Haag
Anna van Saksenlaan 51 | 2593 HW Den Haag
+31 70 349 44 50
info@dans.knaw.nl | dans.knaw.nl

DANS is een instituut van KNAW en NWO



Door data gedreven